

DOCUMENT RESUME

ED 213 768 .

TM 820 218

AUTHOR Jones, Phil; Ligon, Glynn D.
TITLE Preparing Students for Standardized Testing: A Literature Review. Publication No. 81.61.
INSTITUTION Austin Independent School District, Tex. Office of Research and Evaluation.
PUB DATE [81]
NOTE 9p.
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; Drills (Practice); *Literature Reviews; Predictor Variables; Pretests Posttests; Scores; Standardized Tests; *Test Coaching; *Test Wiseness

ABSTRACT

Three variables which affect students' achievement test scores and which can be manipulated prior to test administration are discussed: testwiseness, practice tests, and test practice. Findings were presented from literature reviews. Regarding testwiseness, (1) it can be described, measured, and taught, therefore, testwiseness exists; (2) testwiseness is only mildly related to general intelligence as measured by group intelligence tests and is probably a network of specific skills, not a general cognitive ability; (3) testwiseness increases with maturity and is unrelated to sex; (4) although testwiseness skills can be taught to students of all ages, using a variety of techniques, the affects of such instruction do not last long; (5) the differential effects of testwiseness instruction for students at various points along the testwiseness continuum is unknown and may vary with the type of testwiseness skill under consideration; and, (6) for some groups of students, testwiseness instruction improves the reliability and predictive power of certain tests. No research was found on practice tests' effects on student performance or on the reliability or predictive power of the associated test. Regarding test practice, it was found that taking one IQ test once for practice will improve IQ scores on a subsequent test given no more than two months later, and practice effects may last as long as ten months if the posttest is the same as or a parallel form to the pretest. (Author/GK)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED213768

Preparing Students for Standardized Testing: A Literature Review

PHIL JONES
 GLYNN D. LIGON, Ph.D.
 AUSTIN INDEPENDENT SCHOOL DISTRICT, Austin, Texas

U.S. DEPARTMENT OF EDUCATION
 NATIONAL INSTITUTE OF EDUCATION
 EDUCATIONAL RESOURCES INFORMATION
 CENTER (ERIC)

☒ This document has been reproduced as
 received from the person or organization
 originating it.
☐ Minor changes have been made to improve
 reproduction quality.

* Points of view or opinions stated in this docu-
 ment do not necessarily represent official NIE
 position or policy.

This literature review focuses on three variables which affect students' achievement test scores and which can be manipulated prior to test administration. They are (1) testwiseness, (2) practice tests, and (3) test practice.

Testwiseness

The most commonly used definition of testwiseness is "a subject's capacity to utilize the characteristics and formats of the test and/or the test-taking situation to receive a high score. Testwiseness is logically independent of the examinee's knowledge of the subject matter for which the items are supposedly measures." (Millman, Bishop, and Ebel, 1965, p. 707) Testwiseness affects scores on standardized tests, and the variance caused by testwiseness becomes error variance, i.e., not due to differences in the trait the test seeks to measure.

The construct of testwiseness has been documented well in psychometric literature. Slakter, Koehler, and Hampton (1970a), Diamond and Evans (1972), and Flynn and Anderson (1977) devised instruments to measure testwiseness. Their instruments for measuring testwiseness consist of nonsense items, that is, items based on fictitious material, having no correct answer. However, one option on each item was designated as the "test-wise" response because of the cues contained in the item. For example,

The Davis Act of the 20th Century...

- provided money for schools.
- struck down an earlier law.
- prohibited the manufacture, sale, transportation, or use of several specific drugs that were being used for illegal puposes.
- gave a raise to government employees.

"PERMISSION TO REPRODUCE THIS
 MATERIAL HAS BEEN GRANTED BY

F. Holley

TO THE EDUCATIONAL RESOURCES
 INFORMATION CENTER (ERIC)."

Response c is designated as the test-wise response because it is much longer and more carefully worded than the other three. In the absence of knowledge, the test-wise student would choose response c.

Other cues in the nonsense items (Diamond and Evans, 1972, p. 147) include...

- . resemblances between stem and correct option,
- . grammatical consistency between stem and correct option,
- . specific determiners (words like always and never), and
- . options which imply the correctness of each other.

All of these secondary cues could be used by the test-wise student to spot the correct option and/or eliminate incorrect options.

TM 820 2/8

If students merely guessed, they would average one test-wise response for every four items on the four-choice nonsense test. However, the students in each of the three studies--sixth graders, high school seniors, and college students--chose the test-wise response on more than half the items on the average. This is more than twice the expected success rate from guessing. The students must have been responding to secondary cues.

Further evidence for the existence of testwiseness lies in two facts.

- (1) Students can describe testwiseness (Millman, et al., 1965; and Diamond and Evans, 1972).
- (2) Testwiseness can be taught (see below).

Testwiseness apparently bears only a small relationship to intelligence, as measured by group intelligence tests. Diamond and Evans (1972) found that testwiseness for his sixth-grade subjects was moderately correlated ($r = .45$) to intelligence as measured by the Lorge-Thorndike Intelligence Test. However they concluded that testwiseness is cue specific, not a general cognitive ability. Kreit (1967) found that IQ level among his third-grade subjects was not related to gains in intelligence test scores due to practice. Flynn and Anderson (1977) found no relationship between testwiseness and IQ as measured by the Thurstone Test of Mental Alertness. They also found that students' levels of testwiseness did not influence their final grade in a course. Dunn and Goldstein (1959) found that secondary cue recognition took place equally at all levels of intelligence. Jongsma and Warshauer (1975, p. 14) in reviewing the research on testwiseness, concluded that, "The ability to use test-taking strategies is not related to general intelligence as measured by group intelligence tests," and testwiseness is not a general trait, but "a network of specific and independent skills."

Research by Slakter, Koehler, and Hampton (1970b) and by Crehan (1974) indicates that testwiseness increases with age or grade level, but is unrelated to sex. Research on the relationship between testwiseness and socioeconomic status is scant and inconclusive (Jongsma and Warshauer, 1975).

Testwiseness can be taught to students of all ages, using a variety of techniques. Oakland (1972) taught testwiseness skills to Head Start children by means of six hours of instruction and practice, spread out over a six-week period. Callenbach (1973) found that instruction and practice improved testwiseness skills in second-graders. Jongsma and Warshauer (1975) taught testwiseness to fifth graders by means of a 45-minute reading and discussion unit. Wahlstrom and Boersma (1968) taught testwiseness to ninth graders by lecture and discussion lasting less than two hours. Flynn and Anderson (1977) taught testwiseness to college freshmen by means of taped lectures. Moore (1971) found that a booklet familiarizing college students with analogy question formats and types of analogies led to improved performance on an analogies test.

In summarizing the research literature, Jongsma and Warshauer (1975) conclude that,

Testwiseness can be effectively taught to students of all ages from preschool through adult. Although the results have not always been statistically significant, the gains nearly always favor the instructional group. (p. 18)

Most of the instructional units used in research studies have been brief, no more than one or two hours in length. They have taken various forms, including lecture, self-paced reading, discussion, and programmed instruction. Students can acquire considerable test-taking skills rather quickly. On the other hand, there is ample evidence in the studies cited above that the positive effects of testwiseness instruction do not last very long, probably no more than a month or two. This may be due to forgetting, or it may be due to constant improvement in testwiseness skills by all students as they mature. That students' testwiseness skills may increase as a function of maturation is especially suggested by the Slakter et al., (1970b) and Crehan (1974) studies cited above. Whatever the reason, testwiseness instruction probably needs to be repeated before each test administration to maximize its impact.

Should test-taking skills be taught in school? Experts on all sides say that they should. Researchers such as those we have already discussed favor teaching testwiseness skills in order to reduce error in test scores. Educators, such as Rudman (1976), Erickson (1972), and Downey (1977) are concerned that a lack of test-taking skills will adversely affect the predictive and diagnostic validity of test scores for individual students. Research by Ginther (1978) lends credence to both cases by showing that testwiseness training can favorably affect both the reliability and predictive power of some tests for some groups of students.

As to the predictive and diagnostic validity of test scores for practitioners' purposes, this is a difficult question. What effect would the teaching of testwiseness skills have on the interpretation of the scores? Standardized tests are norm referenced, by definition. How prevalent is the teaching of testwiseness skills to students in the norming sample? We may assume that some proportion of students in the norming sample received instruction in testwiseness skills shortly before they took the test. But how large a proportion? If the proportion is small, then by teaching all our students testwiseness skills, we can raise the percentile scores and grade equivalents of our local students. On the other hand, if the proportion is large, then we underrate our own students and instructional staff if we do not teach testwiseness skills. Furthermore, if the frequent teaching of testwiseness skills to all students becomes standard educational practice, then the next time new tests are normed, the norms may be more demanding since the test-taking skills of the average student would have been upgraded.

Now let us examine the issue of reducing error in test scores. The researchers cited above generally agree that instruction in test-taking skills would reduce such error. However, is this necessarily the case?

Observed test scores can be viewed as the sum of the student's true score on the one hand and some amount of measurement error on the other hand. Measurement error can be caused by many factors including variations in test construction, test-taking conditions, and testwiseness. The assumption by the researchers seems to be that teaching testwiseness in the schools will yield observed scores closer to the true scores by reducing variations in testwiseness as a result of upgrading the skills of the relatively test-naïve subjects.

There are, however, indications in the literature, such as in the Flynn and Anderson (1977) study, that testwiseness instruction may benefit all students equally, regardless of their initial level of testwiseness skills. To the extent that this is true, it would negate the hope that variations in testwiseness could be reduced by teaching all students testwiseness skills. The range of variation in testwiseness skills might remain the same. While the test-taking skills of the relatively test-naïve students would be upgraded

through instruction, so would the skills of the highly test-wise students. It is conceivable that students might have such a high degree of test-taking skill that they are able to "outsmart" the test and receive an observed score that overestimates the students' abilities. This is certainly measurement error, and Flynn and Anderson's finding suggests that this type of error would be exacerbated by teaching testwiseness to all students. Thus, under certain circumstances, instruction in testwiseness may actually increase measurement error in the test scores of some students.

At this point, a distinction is helpful between two types of testwiseness skills. First, basic testwiseness skills include marking answers properly, using available time efficiently, following directions, etc. Basic testwiseness skills help reduce the error in test scores. Second, advanced testwiseness skills include choosing the grammatically correct alternative, avoiding options with "always" and "never," etc. Advanced skills should seldom be advantageous on a well-constructed, standardized test. Thus, the actual chance of impacting error variance on a major standardized test by teaching advanced testwiseness skills appears to be small.

Future research could make a significant contribution to educational practitioners by determining the comparative effects of testwiseness instruction. Do the relatively test-wise always gain as much from testwiseness instruction as the test-naive? If not, under what conditions would the test-naive students gain more than the test-wise? Possibly, Flynn and Anderson's finding that all students benefited equally from testwiseness instruction reflects the fact that advanced testwiseness skills were taught in that study. If only basic strategies had been taught, differential gains might have been observed.

Teaching testwiseness skills only to those students who are deficient in such skills would address all the difficulties outlined above. This strategy would reduce the range of testwiseness skills, upgrade the skills of the test-naive without producing students who could outsmart the test, and have a less troublesome effect on the norms than would the strategy of teaching testwiseness skills to all students.

In a public school setting, however, it may be difficult to put this strategy into practice. In the first place, it would be expensive to target the test-naive students. New measures would have to be designed, normed, and administered, and criterion scores established in order to decide who needs testwiseness skills and who does not. In the second place, it is often difficult for public schools to convince parents that it is equitable to offer certain educational services and resources only to selected students and not to all.

Assuming that testwiseness is to be taught, what specifically should be taught, and how? Millman, et al., (1965) provide a comprehensive outline of the skills and strategies involved in testwiseness. There is no shortage of suggestions on what to teach and how (Rudman, 1976; Erickson, 1972; and Downey, 1977), but a single approach will not work for all grade levels. Students gain in testwiseness and in their ability to grasp testwiseness strategies as they mature and gain experience in test-taking. For the purposes of this review, Jongsma's and Warshauer's summation may be best (p. 39).

Teach those aspects of testwiseness that appear appropriate for your age group. For younger students this might involve such features as following directions, completing answer sheets, and guessing. Older students might profit from more sophisticated skills such as using secondary cues or practicing with particular formats.

Use an instructional format that you believe is effective...There is no research to support one type of instruction over another... Providing concrete examples to illustrate test-taking cues would be advisable.

Repeat instruction in testwiseness...prior to subsequent standardized test administrations.

Oakland's (1972) research suggests that very young students may need to be taught basic concepts, such as right, left, up, down, opposite, most like, same, and different (p. 357).

Future research on testwiseness can make the best contribution to both theory and practice by devoting itself to the question of how best to reduce the measurement error caused by individual differences in test-taking skills. Also, the relationship between testwiseness and socioeconomic status is open for investigation. Preliminary indications (Jongsma and Warshauer, 1975) are that SES may interact with testwiseness instruction in improving test-taking skills.

Practice Tests

A distinction needs to be made between the terms "practice tests," and "test practice." The term practice tests, as used here, refers to packets of exercises for students, designed by test-makers or school personnel, which are much shorter in length than the standardized tests to which they relate, and with items typically much easier than the items on the actual test.

Practice tests are usually designed to

- (1) familiarize students with the visual format of the test,
- (2) familiarize students with the terminology used in the test instructions,
- (3) provide students with practice in using separate answer sheets, and
- (4) ease students' fear of the test.

The authors of this paper were unable to find any research bearing on practice tests and their effects on student performance outcomes or on the reliability or predictive power of the associated test.

Test Practice

Let us define "test practice" as the use of full-length, standardized tests as practice exercises prior to the administration of a standardized test for official purposes.

Especially where intelligence and aptitude tests are concerned, there is ample research related to the effects of test practice on test scores. The findings regarding intelligence tests are remarkably consistent. Three separate researchers (Peel, 1952; Kreit, 1967; and Eichelberger, 1970) have found that one and only one test practice session, within one month of the official test, will improve IQ scores. Any further test practice is useless, they conclude, because a second or third practice session does not affect IQ scores, regardless of the time interval between test practice and the official test. If the interval is more than two months, the official test scores probably will be unaffected by test practice. Nero (1976) found with an aptitude battery that practice effects

lingered for as much as ten months if the posttest was either the same as the pretest or a parallel form. Nero's study further shows that improved performance based on test practice does not result from repetition (at least at 10-month intervals), but results instead from subjects' being familiar with the specific format of the test.

Summary

The following statements summarize findings from the literature reviewed.

Testwiseness

- (1) Testwiseness can be described, measured, and taught; therefore, testwiseness exists.
- (2) Testwiseness is only mildly related to general intelligence as measured by group intelligence tests and is probably a network of specific skills, not a general cognitive ability.
- (3) Testwiseness increases with maturity and is unrelated to sex.
- (4) Although testwiseness skills can be taught to students of all ages, using a variety of techniques, the affects of such instruction do not last long. Therefore, such instruction must take place within two months of the actual test administration in order to have its maximum impact.
- (5) The differential effects of testwiseness instruction for students at various points along the testwiseness continuum is unknown and may vary with the type of testwiseness skill (basic or advanced) under consideration.
- (6) For some groups of students testwiseness instruction improves the reliability and predictive power of certain tests.

Practice Tests

No research was found on practice tests' effects on student performance or on the reliability or predictive power of the associated test.

Test Practice

- (1) Taking one IQ test once for practice will improve IQ scores on a subsequent test given no more than two months later. Any more test practice than this produces no more improvement.
- (2) Practice effects may last as long as ten months if the posttest is the same as or a parallel form to the pretest.

References

- Callenbach, C. The effects of instruction and practice in content-independent test-taking techniques upon the standardized reading test scores of selected second-grade students. Journal of Educational Measurement, 1973, 10, 25-30.
- Crehan, K. D. Longitudinal studies of testwiseness. Journal of Educational Measurement, 1974, 11, 209-212.
- Diamond, J., and Evans, W. An investigation of the cognitive correlates of testwiseness. Journal of Educational Measurement, 1972, 9, 145-150.
- Downey, G. W. Is it time we started teaching children how to take tests? American School Board Journal, 1977, 164, 27-31.
- Dunn, T. F., and Goldstein, L. G. Test difficulty, validity, and reliability as functions of selected multiple choice item construction principles. Educational and Psychological Measurement, 1959, 19, 171-179.
- Eichelberger, R. T. Practice effects of repeated I. Q. testing and the relationship between I. Q. change scores and selected individual characteristics. (Doctoral dissertation, Southern Illinois University, 1970). Dissertation Abstracts International, 1971, 31, 5196A.
- Erickson, M. E. Test sophistication: an important consideration. Journal of Reading, 1972, 16, 140-144.
- Flynn, J. T., and Anderson, D. E. The effects of test item cue sensitivity on I. Q. and achievement test performance. Educational Research Quarterly, 1977, 2, 32-39.
- Ginther, Joan R. Pretraining Chicano students before administration of a mathematics predictor test. Journal for Research in Mathematics Education, 1978, 9, 118-125.
- Jongsma, E. A., and Warshauer, E. The effects of instruction in test-taking skills upon student performance on standardized achievement tests. New Orleans, La., New Orleans University, 1975. (ERIC Document Reproduction Service No. ED 114 408).
- Kreit, L. H. The effects of practice on the acquisition of test-taking skills by a selected group of third-grade pupils (Doctoral dissertation, Indiana University, 1967). Dissertation Abstracts International, 1968, 28, 3511A.
- Millman, J., Bishop, C. H., and Ebel, R. An analysis of testwiseness. Educational and Psychological Measurement, 1965, 25, 707-726.
- Moore, J. C. Testwiseness and analogy test performance. Measurement and Evaluation in Guidance, 1971, 3, 198-202.

- Nero, B. The effects of general practice, specific practice, and item familiarization on change in aptitude test scores. Measurement and Evaluation in Guidance, 1976, 9, 16-20.
- Oakland, T. The effects of testwiseness materials on standardized test performance of preschool disadvantaged children. Journal of School Psychology, 1972, 10, 355-360.
- Peel, E. A. Practice effects between three consecutive tests of intelligence. British Journal of Educational Psychology, 1952, 22, 196-199.
- Rudman, M. K. A few hints for teaching students the skills needed for taking standardized tests. Learning, 1976, 4, 76-82.
- Slakter, M. J., Koehler, R. A., and Hampton, S. H. Learning testwiseness by programmed texts. Journal of Educational Measurement, 1970, 7, 247-254. (a)
- Slakter, M. J., Koehler, R. A., and Hampton, S. H. Grade level, sex, and selected aspects of testwiseness. Journal of Educational Measurement. 1970, 7, 119-122. (b)
- Vernon, P. E. Symposium on the effects of coaching and practice in intelligence tests. British Journal of Educational Psychology, 1954, 24, 57-63.
- Wahlstrom, M., and Boersma, P. J. The influence of test-wiseness upon achievement. Educational and Psychological Measurement, 1968, 28, 413-420.